2018

Measurement of the Propensity to Trust Automation

Sarah Ann Jessup
Wright State University

MEASUREMENT OF THE PROPENSITY TO TRUST AUTOMATION

A thesis submitted in partial fulfillment of the
requirements for the degree of
Master of Science

By

SARAH ANN JESSUP
B.S., Wright State University, 2015

2018
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

NOVEMBER 19, 2018

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Sarah Ann Jessup ENTITLED Measurement of the Propensity to Trust Automation BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Tamera R. Schneider, Ph.D.
Thesis Director

Scott N. J. Wattamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

Committee on
Final Examination

Gary N. Burns, Ph.D.

Gene M. Alarcon, Ph.D.

Barry Milligan, Ph.D.
Interim Dean of the Graduate School

ABSTRACT

Jessup, Sarah Ann. M.S. Department of Psychology, Wright State University, 2018.
Measurement of the Propensity to Trust Automation.

Few studies have examined how propensity to trust in automation influences trust behaviors, those which indicate users are relying on automation. Of the published studies, there are inconsistencies in how propensity to trust automation is conceptualized and thus measured. Research on attitudes and intentions has discerned that reliability and validity of measures can be increased by using more direct and specific language, which reduces ambiguity and increases the ability to predict behavior. This study examined how traditional measures of propensity to trust automation could be adapted to predict whether automation is deemed as trustworthy (perceived trustworthiness) and whether people behave in a trusting manner when interacting with automation (behavioral trust). Participants (N = 55) completed three propensity to trust in automation surveys including Propensity to Trust in Technology, an adapted version, and the Complacency-Potential Rating Scale. The Propensity to Trust in Technology scale was adapted by replacing "technology" with "automated agent" as the referent. Participants played a modified investor/dictator game, where people teamed with a NAO robot. Betting behaviors were used to measure behavioral trust.

This study demonstrated that compared to generally-worded measures, more specifically-worded measures of propensity to trust automation are more reliable and better predictors of perceived trustworthiness and behavioral trust. An adapted propensity to trust technology scale was the only significant predictor of both perceived trustworthiness of the automation and the trusting behaviors of participants. By decreasing the ambiguity of the referent in the adapted propensity to trust automation scale, the reliability and predictive validity was increased.

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

As systems, tasks, and machines become increasingly automated, researchers must understand how people respond to such automation. Automation refers to technology that allows for data to be collected, selected, computed, and analyzed (Lee & See, 2004). Automation helps users control processes in a systematic way, make decisions, decrease cognitive workload, and reduce human error (Hoff & Bashir, 2015; Huang & Bashir, 2017; Miller, 2018). For example, medical professionals rely on robots for precision, dexterity, and endurance. Increasingly, car companies are producing automated vehicles, with the goal of improving safety. Aviation has used automation for decades, and it is now possible to land a plane using an iPad (Adams, 2014). These are only a few domains incorporating automation into tasks that were once performed entirely by humans. However, automation is irrelevant if humans do not use these automated systems because they do not trust the systems. As more systems become automated, learning how humans trust, and how they trust automation in particular, and the factors that affect the use of automation effectively is important.

Propensity to trust is the general tendency and willingness of one person to trust another person (Mayer, Davis, & Schoorman, 1995). Research has examined the relationship between propensity to trust and trust in interpersonal situations, and researchers have found that the propensity to trust another person predicts initial trustworthiness, the perception of how trustworthy another person is, toward the other person (Alarcon, Lyons, & Christensen, 2016; Colquitt, Scott, & LePine, 2007; Gill, Boies, Finegan, & McNally, 2005). Relatively recently, researchers have adapted and extended the interpersonal trust literature to include trust in automation. However, as discussed below, trust in another human and trust in an automated system might not be parallel. There are few studies that have examined how the propensity to trust in automation influences behaviors that suggest the user is relying on that automation (behavioral trust) in human-automation contexts. Of the studies that are published, there are inconsistencies in how propensity to trust automation is conceptualized and thus measured. Researchers have used one scale to measure automation complacency, propensity to trust, and automation expectancy (Merritt & Ilgen, 2008; Pop, Shrewsbury, & Durso 2015; Singh, Molloy, & Parasuraman, 1993), and some measures of propensity to trust have demonstrated low reliabilities (Merritt & Ilgen, 2008; Schneider et al., 2017). If researchers know how to measure personality and dispositions well, then they are better able predict behavior (Funder & Ozer, 2010; Wiggins, 1973). The purpose of this study was to examine how traditional measures of propensity to trust automation could be adapted to predict the perceived trustworthiness of automation and behavioral trust.

Evolution of the Trust Literature

Researchers have studied trust for several decades, refining the way they have conceptualized and defined trust over time. Initially, trust was described as an attitude that a person develops very early in life (Erickson, 1950). This trusting attitude is the belief a person has that other people are generally trustful and that they themselves are trustworthy. A decade later, trust was defined as a dispositional trait as well as situational circumstance (Deutsch, 1960). In other words, trust was both a characteristic of a person and dependent on the situation or context. Trust has been defined as an expectancy that others can be relied upon and stay true to their word (Rotter, 1967), an intention or a willingness to be vulnerable (Mayer et al., 1995), and as a psychological state (Rousseau, Sitkin, Burt, & Camerer, 1998). These ways of thinking about trust shift the focus onto people trusting other people. Research on people trusting automation has focused on trust as an attitude that automation will help an individual achieve her goal under novel and uncertain situations (Lee & See, 2004). Trust research has come full circle, from describing trust of another as an attitude by Erikson (1950) to Lee and See (2004) also suggesting that trust is an attitude, but with a focus on trusting automation. The ways of thinking about trust have changed over time.

Trust Models

Interpersonal Trust. In an effort to integrate and refine the many definitions of trust, as well as the trust process, Mayer et al. (1995) proposed the dominant interpersonal trust model. This model includes a trustee or referent of trust, and a trustor or the person or group engaging in trusting intentions or behaviors. The model describes how characteristics of the trustee and the disposition of the trustor both influence trust in different ways. The outcome of the trust process proposed by Mayer et al. (1995) is "risk taking in relationship" (p. 725). The authors separate general risk-taking behaviors (e.g. gambling or skydiving) and taking risk in interpersonal relationships. The latter is specified as actual trusting behavior.

The characteristics of the trustee are those factors that help determine the perceived trustworthiness of the trustee, as determined by the trustor. Mayer and colleagues suggest that there are three components or characteristics of trustees that go into determining whether the trustee will be perceived as trustworthy by the trustor. These factors are ability, benevolence, and integrity. The first factor, ability, is the level of knowledge or the skills the trustee possesses, which is domain-specific. Even though a trustee has the ability to carry out one task (i.e. financial investing), she might not have the skill set to accomplish a different task, such as emergency medicine. The second factor, benevolence, denotes that the trustor believes the trustee has her best interests in mind when making decisions or engaging in behaviors. Using the same example as above, if a financial investor is benevolent, then she will invest money in securities that benefits her clients, even though these investments might not earn much commission for herself. Integrity, the third factor, is the trustor's perception that the trustee will behave ethically and lawfully. When engaging in a financial investment deal, the client trusts the financial investor to invest the money in an agreed upon security and to not launder the money into the financial investor's personal bank account. Both these characteristics of the trustee and the propensity of the trustor to trust influence what transpires in a given context in distinct and significant ways.

Propensity to trust is a general willingness and tendency to trust other people (Mayer et al., 1995). Propensity to trust is a relatively stable disposition that is developed over time. As people engage with others in their lives, these interactions, positive and negative, will begin to shape their beliefs about other people. The experiences that people have personally with others, events that people witness others experiencing, and the experiences that individuals learn about through media and society all shape how people develop their own predispositions to trust (Van Lange, 2015). Before a person ever meets or interacts with another, the trustor will have some internal tendency to higher or lower levels to trust another, which will influence initial intentions to be vulnerable to another. People differ in their levels of propensity to trust. When meeting new people, propensity to trust will affect how likely it is that a person (the trustor) will trust another person (the trustee) before the trustor knows anything about the trustee. For example, if an individual's car broke down on the side of the road and a stranger stopped and offered to give the individual a ride, the individual's predisposition to trust will influence if he will accept the ride or not. A person who has a high propensity to trust compared to someone with a low propensity to trust would be more willing to accept the ride with the stranger. Propensity to trust predicts trusting behaviors in novel situations when people rely on their past experiences and general trusting or distrusting dispositions (Rotter, 1980). McKnight, Cummings, and Chervany (1998) suggested that propensity to trust will lead to trusting beliefs especially in novel or ambiguous situations. As more information about a trustee becomes available, propensity to trust is less likely to be a significant predictor of trust (Jones & Shah, 2016). The trustor will begin to rely on other characteristics of the trustee such as whether the trustee adheres to her word, reciprocates trust in subsequent interactions, or acts in a trustworthy manner. The feedback that the trustor receives from interactions with the trustee will be more of predictor of trust than his own propensity to trust.

Propensity to trust and perceived trustworthiness are antecedents of trust that should influence an action or behavior that denotes making oneself more or less vulnerable to another person or party. Mayer et al.'s (1995) model describes factors influencing trust between humans and does not consider that humans also interact with automation and trust may influence those interactions. Lee and See (2004) adapted prior models to specifically focus on the factors that influence whether people will trust automation.

Trust in Automation. Lee and See (2004) have suggested that there are factors related to the trustworthiness of automation. Similar to the three factors of perceived trustworthiness found in the interpersonal trust model by Mayer and colleagues, Lee and See propose three factors of perceived trustworthiness of automation. These three factors relate to the interpersonal trust factors of ability, benevolence, and integrity and they are performance, purpose, and process (Lee & Moray, 1992), respectively. Performance describes what the automation does and what the automated system is able to do or capable of doing in the future. Purpose is the second factor, which denotes why the automation is used by people and whether it performs those functions. The third factor, process, which corresponds to how well the automation functions such as whether it is reliable or not. Process allows for the user to see what functions the automation is performing. Performance, purpose, and process are characteristics of automation that can be viewed as variable by human users. Their model also accounts for individual differences of the user.

Lee and See (2004) described the predisposition to trust in much the same way as

Mayer et al. (1995) described the propensity to trust. Predisposition to trust is characterized as an individual difference that influences initial trust intentions prior to any specific interaction or knowledge about the automation. A person's predisposition to trust is comprised of past interactions with automation and influences how new information about subsequent interactions will be interpreted. These characteristics are inherent to the individual and not directly related to characteristics of the automation. Propensity to trust rather than predisposition to trust will be used hereafter to refer to the tendency to trust automation, because it is more common in the human-human and the human-automation trust literature (e.g. Colquit et al., 2007; Schaefer, Chen, Szalma, & Hancock, 2016). There are other factors that influence trust in automation such as culture, display design, and environment, to name a few, that are beyond the scope of this paper. The trust in automation model also accounts for the consequences of trusting, such as reliance on automation, which is defined as trusting behaviors demonstrated by using the automation. Just as the human-human trust literature has evolved, the human-automation trust literature is evolving, which is clearly demonstrated by the changes in measuring this construct.

Evolution of Measuring Propensity to Trust

Interpersonal Propensity to Trust. Researchers have measured the propensity to trust using various scales. According to a meta-analysis on trust (Colquitt et al., 2007), researchers have measured propensity to trust using a Faith-in-People scale (Rosenberg, 1956), an Interpersonal Trust scale (Rotter, 1967), the trust facet of the Agreeableness scale on the NEO PI-R (Costa & McCrae, 1992), and a Propensity to Trust scale (Mayer

& Davis, 1999). Some of the items included in these measures reference specific referents such as, “Most salespeople are honest in describing their products” (Mayer & Davis, 1999), while other measures use more general items that apply to society as a whole, “Trust others” (Costa & McCrae, 1992).

In previous research on interpersonal trust, the propensity to trust has predicted both perceived trustworthiness and behavioral trust. Alarcon, Lyons, and Christensen (2016) used a modified version of the Prisoner’s Dilemma task to explore the relationship between propensity to trust and perceived trustworthiness in both familiar and unfamiliar participant dyads. The propensity to trust was measured using Mayer and Davis’ (1999) Propensity to Trust scale and perceived trustworthiness was measured using a modified version of the Trustworthiness scale (Mayer & Davis, 1999). Alarcon et al.’s results indicated that when the participants did not know one another, their propensity to trust was significantly, positively related to perceived trustworthiness of their partners. In related research, Evans and Revelle (2008) found that participants’ propensity to trust (measured by a scale the researchers created for their study) predicted their behavioral trust while playing the Investment game. More specifically, individuals with higher propensity to trust compared to individuals with lower propensity to trust invested more money in their partners over the course of the experiment.

Propensity to Trust Automation. Just as researchers have defined and measured the propensity to trust between humans, researchers have pursued similar efforts for humans and their work with automation. Research on the measurement of the propensity to trust automation is far less researched than the propensity to trust humans. As such, there are few scales that have been developed to measure this construct. There are two measurements of the propensity to trust automation that will be focused on in this paper. The two measures that will be discussed are the Complacency-Potential Rating Scale (CPRS; Singh et al., 1993) and the Propensity to Trust in Technology scale (PTT; Schneider et al., 2017). The CPRS is one measure that has been used in a variety of studies. The PTT is a new scale with limited use. These two scales were chosen because there has been empirical evidence in the trust in automation literature supporting the use of the CPRS. However, the CPRS is unique in that the items reference several specific types of automation (e.g., ATMs, cruise control, automated devices involved in aviation). Conversely, the PTT contains items that are more general and contains only one referent, technology. To the author's knowledge, there is no study in the trust in automation literature that has empirically compared and contrasted the two types of scales.

***Complacency-Potential Rating Scale*.** One way researchers have measured the tendency for humans to trust automation is by using the CPRS (Singh et al., 1993). The authors defined complacency as a state of low suspicion. When automation is performing as it should, and the users are satisfied with the performance of the automation, users might become less alert and attend less to the automated system, which can be characterized as complacency. Guided by the premature cognitive commitment theory (Langer, 1989), the researchers proposed that either favorable or unfavorable attitudes are created from initial encounters with automation. These attitudes become reinforced during subsequent interactions. The purpose of the CPRS was to measure attitudes that people may have towards automation. More specifically, the authors wanted to examine if peoples' attitudes determine if they are more or less willing to rely on automated systems.

Researchers examined three domains that were rich with human-automation interactions to create items for the CPRS (Singh et al., 1993). The domains that provided data included Aviation Safety Reporting System reports, subject-matter experts (SMEs), and prior computer-attitude scales. From delving into these three data-rich resources, the authors created 100 items to measure attitudes towards automation. The items were worded to convey attitudes toward using automation as either a benefit or as a risk. Four SMEs chose 20 of these 100 statements with the highest face validity that assessed automation complacency to be included in the CPRS. Four of the items were filler items. A cross-sectional study (N = 139) was conducted to examine the psychometric properties of the CPRS (Singh et al., 1993). The items on the scale were highly intercorrelated (r > .98). Factor analysis revealed a total of five factors: 1) general automation, 2) confidence-related complacency, 3) reliance-related complacency, 4) trust-related complacency, and 5) safely-related complacency. The five factors accounted for 53% of the variance in the items, with an overall Cronbach's alpha of .90. A follow-up study examined test-retest reliability. The researchers mailed surveys to the original 139 participants, with 44 participants returning the surveys. Internal consistency (α = .90) and three months later test-retest reliably (α = .87) were high. The CPRS was not significantly correlated with age, education, attitudes towards computer use, or computer experience. This scale was one of the first to examine trust in automation by assessing attitudes toward complacency. However, the CPRS was not used to predict actual complacency behaviors.

Other researchers have used the CPRS to predict trust behaviors during human-automation interactions. Merritt and Ilgen (2008) used 12 items from the CPRS to measure propensity to trust machines (see Appendix A). The purpose of the study was to examine trust in automation use. The task that the authors used to measure automation use was a luggage screening task similar to what would be used in an airport. The automation in the study was an x-ray screening task equipped with an automated weapon detector to examine luggage for potential threats, such as guns and knives. An important dependent variable was participants' use of automation, which was measured by the number of bags that were screened using the automation. Participants had to activate the automation if they wanted to use it during the task. Prior to the start of the luggage screening task, participants were exposed to automated weapons detector for one minute. After the interaction, participants were asked to answer self-report questions about the characteristics of the automation using a trust scale created for the study to measure pre-task perceived trustworthiness of the automation. An example item is, "I have confidence in the advice given by the AWD [automated weapons detector]". Post-task trustworthiness was measured after 20 minutes of interaction with the automation. Results indicated that propensity to trust machines was significantly positively related to initial trustworthiness ($r = .23$, $p < .05$). Propensity to trust was not significantly related to post-task trustworthiness ($r = -.04$, $p > .05$), or the actual use of the automation ($r =$

.14, $p > .05$). One reason for some of these null findings could be that the reliability of the CPRS was low ($\alpha = .61$), which diverges from previous research. Reasons for this low reliability might be that some of the items reference technology that was outdated near 2008, which is when this study was published. For example, one item on the scale references using a VCR, which was less relevant in 2008 compared to 1993. Another possibility for the low reliability is that Merritt and Ilgen used a shortened version of the scale that might not measure the construct as completely as the 16 items that measured complacency in 1993. Merritt and Ilgen's study examined how propensity to trust influenced trustworthiness of the x-ray automation to screen luggage, and how it influenced using that automation, with the main finding that propensity to trust machines was related to perceived trustworthiness of the automated system.

Propensity to Trust in Technology. A recently developed measure of propensity to trust automation is the Propensity to Trust in Technology (PTT; Schneider et al., 2017). The scale was designed to measure the general tendency to trust in technology. The creation of the scale was driven by theory related to trust in automation and the factors that are purported to influence trust (Lee & See, 2004). The six items (see Appendix B) were designed to measure stable characteristics in individuals, attitudes towards technology, and whether people were likely to collaborate with technology. An example item is, "I think it's a good idea to rely on technology for help."

To empirically test the measure, Schneider et al. (2017) used the scale in a study examining human-automation collaboration. Prior to the start of the task, participants ($N = 44$) completed the PTT ($\alpha = .64$). Participants were then introduced to a virtual spaceship environment through which the human-automation task would be accomplished. The participant, a virtual robot teammate (CEP), and captain of the spaceship were presented as avatars within a virtual world. The captain explained that the participant would be working with a robot and that the robot had previous experience with the task. Then, the captain informed the participant that an emergency landing to the moon was inevitable and to make a list of 10 items they would need on the moon. The participant was told she could work with CEP and that both of their rankings would be evaluated, relative to other teams, by the captain. The participant and CEP independently ranked 10 items they would take to the moon with them in order of importance, 1 (most important) to 10 (least important). Unbeknownst to participants, CEP's rankings remained the same across participants. After submitting their rankings, the participants could compare their rankings with CEP's. In this comparison, CEP's rankings included the rationale CEP used for the ranking order. The participant could change her rankings and was to submit a final ranked list to the captain. Rankings were calculated as the absolute difference between the participant's initial and final rankings (P-P Rank). The researchers also computed the difference between the participant's final rankings and CEP's rankings (P-CEP Rank). Results indicated that the PTT was marginally related to P-P Rank ($r = .27$, $p = .09$) but not related to P-CEP Rank ($r = .22$, $p = .18$), demonstrating that initial trust levels tended to be related to a change in ranking, but not relative to CEP's ranking. It could be that that as more information about working with CEP across the task and the rationale for CEP's rankings became available, propensity to trust was less of a predictor of behavioral trust (Jones & Shah, 2016).

Participants may have been relying on their perceived trustworthiness of CEP rather than their initial trust in automation tendencies. The participants' perceptions of trustworthiness of CEP might have been a better predictor of trust in the automation over the course of the study. One limitation of this study is that the PTT had low scale reliability. Using more precise language when creating the items so that the referent reflects the referent in the study may increase reliability to an acceptable level, which is .70 or higher (Nunnally, 1978).

General vs Specific Scale Specificity

When creating measurements of attitudes and intentions, using more direct and specific language reduces ambiguity and increases prediction of behavior (Ajzen & Fishbein, 1973). If a measure of intention is abstract or general, the measure will have a lower correlation with the actual behavior than would a more specific or precise measure. For example, if women over the age of 50 are asked whether they intend to get an annual mammogram, most women might say yes, they do intend on getting a mammogram. However, if the question is phrased, "Do you plan to get a mammogram in the next 12 months," most women might respond with "no" compared to how the question was presented previously. This might occur because the latter question provides a more specific time line. The women may begin to think about their insurance coverage, their current health status, and other factors that might make the women more inclined to say no rather than yes. Improving the precision of how the question is framed increases the predictive utility of that question with their actual behavior within the next year (Schneider et al., 2001).

In addition to improving the precision of language that researchers use when creating measures, providing participants with a frame of reference can also improve the reliability and validity of measurement. Frame of reference effect refers to the phenomenon that occurs when participants are given more context in personality tests. When participants are provided with a context to which the question or item applies, then the reliability and validity of the measure is improved (Lievens, De Corte, & Schollaert, 2008). This is because when participants are provided with a noncontextualized item, the variability between-participants increases. When participants are not provided with context each participant might be thinking of how they behave in different situations across participants.

Researchers have examined the use of specific personality tests in the workplace (Bowling & Burns, 2010). These researchers administered a workplace specific personality test to employees. To create the specific workplace measure, the researchers appended the words "at work" to International Item Pool Items that measured the Big 5 personality traits of conscientiousness, extraversion, agreeableness, and emotional stability. The outcome variables that the authors used to predict workplace criteria were comprised of job satisfaction, work frustration, turnover intention, and absenteeism. They found that when employees were administered the workplace specific personality test, this specific measure of personality accounted for more variance in predicted workplace outcomes over and above the original personality measure. When participants were given a narrower context to focus, on the ambiguity about the items was reduced. Consequently, the predictive validity of the measure about the context of the workplace improved. For example, when people are asked to rate their general tendency to "Follow a schedule," they may respond differently depending on whether the context is their home life, work life, or generally how they view themselves. If the context is the same for every item, then people are more likely to answer the items more consistently for that particular context, increasing both the reliability and the predictive validity of the measure for that context (Schmit, Ryan, Stierwalt, & Powell, 1995).

Using more precise language when creating scales that are designed to measure a construct improves the reliability of the measure, as well as the predictive validity of the measure. As such, the language of one of the propensity to trust automation measures in this study will be adapted. The Propensity to Trust in Technology scale will be adapted so that the referent is "automated agent" in place of "technology." Technology refers to broad "practical applications of knowledge" (Merriam-Webster, n.d.) that might be automated, but by definition, technology does not necessarily refer to automation. For example, a light bulb would be considered technology. By using "automated agent" as the referent of trust instead of "technology", ambiguity is expected to be reduced and the context of the items is more specific than technology. The participants in this study were given a more specific frame of reference that, similar to previous research, should improve both reliability and validity of the measure.

Hypotheses

The purpose of the study was to examine measures of the propensity to trust automation, and their ability to predict initial perceived trustworthiness and initial behavioral trust. Based on prior research, the following hypotheses were proposed:

Hypothesis 1: Propensity to trust automation will predict perceived trustworthiness.

Hypothesis 2: Propensity to trust automation will predict behavioral trust.

Hypothesis 3: The adapted propensity to trust in technology scale will better predict perceived trustworthiness compared to the other two propensity to trust automation scales.

Hypothesis 4: The adapted propensity to trust in technology scale will better predict behavioral trust categories compared to the other two propensity to trust automation scales.

II. METHOD

To test hypotheses, correlational, hierarchical multiple regression, and discriminant function analyses were computed. An a priori power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) by selecting a two-tailed bivariate normal model correlation test, estimating a medium effect size ($\rho = .30$), $\alpha = .05$, and power ($1 - \beta$) = .80. Results from the power analysis indicated a sample size of 84 was needed. Additionally, an a priori power analysis was conducted by selecting a linear multiple regression: fixed model, R^2 increase, estimating a medium effect size ($f^2 = .15$), $\alpha = .05$, and power ($1 - \beta$) = .80. Results from the power analysis indicated a sample size of 55 was needed. For discriminant function analysis, the number of predictors should be smaller than the number of participants in the smallest group (Tabachnick & Fidell, 2007). Because the smallest group (small loan amount) contained 12 participants and the number of predictors is three, this criterion was met. Because hierarchical multiple regression and discriminant function analysis were the main focus of this study and used to test Hypotheses 2, 3, and 4, 55 participants were recruited for this study.

Participants

Participants were 55 adults recruited from a Midwestern college. Ages ranged from 18- 41 years[1] (*M* = 24.11 years, *SD* = 5.36 years). Ethnicity ranges were 41.8% Black or African American, 36.4% White, 16.4% Asian, and 5.5% Biracial. Most (64%) were female. 72.7% of participants listed English as their native language, 5.5% were

[1]One participant chose 86 years old, but no participant fit this description and that age may have been a typo and was left blank.

Hindi native speakers, 3.6% Mandarin, and 3.6% Igbo. The remaining 14.4% of participants listed another language that made up less than 2% each of the sample. For level of education, 40% had some college, no degree, 18.2% were college graduates, 18.2% had a post-graduate degree, 14.5% listed high school as their highest level completed, 3.6% grade school, and 3.6% had completed some post-graduate work. Participants were recruited from the Introduction to Psychology (PSY 1010) participant pool, flyers, email, and word of mouth. Participants received compensation in the form of a $30 gift card, as well as cash payment for all money earned during the task. All data were collected in a two-room laboratory on campus. The front room contained four desktop computers separated by partitions, and this is where participation occurred. The back room had two computer stations, with one being the experimenter computer and the NAO robot was hooked up to a laptop computer at the other station.

Task

The task is called Checkmate (Alarcon et al., 2017). It is a computer game played between two players. Checkmate is a modified version of the investor/dictator game (Berg, Dickhaut, and McCabe, 1995). The investor/dictator game provides an opportunity for researchers to study trust between two people. The goal of the game, from the participants' perspective, is to earn money during the course of the game. The investor/dictator game is played between two players. One person is assigned the role of the "investor" and the other person is assigned the role of the "dictator." The experimenter gives the investor an amount of money to use during game play. The investor chooses an amount to send to the dictator. Once the dictator receives the investment, she decides an amount to send back to the investor. This game can be played over one or several rounds. If the investor sends over money to the dictator, then he is showing that he trusts the dictator to send money back. If the dictator sends money back to the investor, she is showing that she is trustworthy. This game is transparent in that both players know how much money the other has total at any given time; there is no ambiguity. Trust is the willingness to vulnerable to others, *regardless of the ability to monitor their actions* (Mayer et al.; 1995). In the investment game, players are essentially able to monitor the actions of their partners, which begs the question of whether trust is being accuracy conceptualized in this type of game. Checkmate was developed so that some ambiguity is present, leading to a more accurate representation of trust. In Checkmate, participants are not able to monitor their partner as much, therefore creating an environment of trust that is more representative of how people trust others outside of a lab setting.

In the current study, the participant was assigned the role of the "banker" (investor in the investment/dictator game) and a robot played the role of the "runner" (dictator in the investment/dictator game). The role of the banker was to loan money to the runner over the course of five rounds. The role of the runner was to collect boxes in a virtual maze over the course of five rounds. The number of boxes collected by the runner reflected performance.

The initial amount of money the banker had in his virtual account was set at $50. The banker loaned money to the runner each round in anticipation of earning interest on his investment. The banker chose one of three options: loan a small amount to the runner

($1-$7), loan a medium amount ($4-$10), or loan a large amount ($7-$13).

The runner chose a risk level for the purpose of potentially increasing the initial loan amount. The risk levels ranged from low (75-150%) to moderate (50-200%) to high (0-300%). The runner could earn more money by choosing a higher risk level, but the runner risked not earning any money at all if her performance was poor. If the runner decided to err on the side of caution and chose a low risk level, the maximum amount of money the runner lost was 25% without collecting any boxes or gained 50% by performing well.

At the beginning of the round, the runner chose a risk level. The runner then promised to return the initial loan and 50% of the earnings to the banker. The banker was notified via a pop-up message which risk level the runner selected, as well as how much of the invested money the runner promised to return. At this point in the round, the banker selected an amount to loan to the runner: Small ($1 - $7), Medium ($4 - $10), or Large ($7 - $13). Money was then transferred into the runner's virtual wallet. The maze-running task begun, and the banker was able to watch a top-down video of the runner's progress. The runner was allotted two minutes to collect as many boxes as possible. After the maze-running task was over, the runner then decided how much money to return to the banker. The banker received a pop-up message of the exact amount of money the runner decided to return.

The steps outlined above were repeated over five rounds. Participants were informed that all money exchanged in the task represented real money. The amount of money the banker had in his/her virtual bank at the end of the session belonged to the banker, and the earnings were paid out in the form of cash, rounded up to the nearest quarter.

Manipulations

Typically, Checkmate (Alarcon et al., 2017) is played between two people, one is the banker and the other is the runner. For this study, the participant was always the banker and the runner was always a Nao robot (see Appendix C). The runner's risk level in the game was set to medium-risk for every round, and the runner always returned the promised amount of money to the banker. All the runner's data, including maze performance, returning of investment to banker, was prerecorded. This level of control allowed a focus on the way that participants trusted the automated partner. However, participants were led to believe they were playing in real time with the robot.

Measures

Propensity to Trust Automation.

Complacency-Potential Rating Scale. The original Complacency-Potential Rating Scale (CPRS; Singh et al., 1993) is a 20-item scale designed to measure attitudes towards automation that could influence the potential for complacency, which includes 4 filler items. The scale has been shown to have high reliability ($\alpha = .90$), high internal consistency ($r > .98$), and high test-retest reliability ($\alpha = .87$). To this researcher's knowledge, the full version of the scale is not published. Singh et al. only provided a few items for each sub-scale, and they did not provide examples of the filler items. The authors only published 14 of the 20 items. As such, a shortened 14-item version was used for the present study (see Appendix A). The 12-item version has been used to in previous research to measure propensity to trust machines (α = .61; Merritt & Ilgen, 2008). Participants were asked to indicate their agreement or disagreement with each of the statements using a 5-point response scale ranging from 1 (strongly disagree) to 5 (strongly agree). Scores were computed by taking an average of all items, after reversing items 7 and 8. An example item from the scale is, "If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because it is more reliable and safer than manual surgery." Because the items on this scale reference specific types of automation such as cruise control and ATMs, the items were not adapted for this study. This scale had poor reliability in this study (α = .67).

Propensity to Trust in Technology. The original Propensity to Trust in Technology (PTT; Schneider et al., 2017) scale has 6-items that measure the general propensity to trust in technology (α = .65). The items on this scale were designed to measure stable characteristics of individuals, attitudes towards technology, and the potential for collaboration with technology (see Appendix B). Participants were asked to indicate their agreement, from 1 (strongly disagree) to 5 (strongly agree), with each of the statements. Item 4 were reversed scored. An example item is, "Technology helps me solve many problems." An adapted version of this scale using *automated agent* as the referent was used for this study (see Appendix D). Using the example above, the item became, "Automated agents help me solve many problems." This scale was reliable in this sample (α = .76). During pilot testing, a few participants asked what an automated agent meant. To decrease ambiguity, a definition of an automated agent was included after the instructions for the survey: "An automated agent can be defined as an entity that runs by computerized algorithms and that interacts with humans."

Perceived Trustworthiness. Perceived trustworthiness was measured using the Trust in Automated Systems (Jian, Bisantz, & Drury, 2000) scale (α = .96; Chancey, Bliss, Proaps, & Madhavan, 2015). This 12-item scale was designed to measure trust in a specific type of system (see Appendix E). The original scale references "the system" as the referent. The scale was modified to indicate the runner was the referent. Only the adapted version was used in this study (see Appendix F). Item 3, "I am suspicious of the system's intent, action, and outputs" was removed because it is a triple loaded item. The 11-item adapted measure was used as an outcome variable to measure self-reported perceived trustworthiness in the runner. Participants were asked to select the option that best described their feeling or impression of the runner during the interaction using a seven-point scale ranging from 1 (not at all) to 7 (extremely). All items were randomized in the study because all reverse-scored items were at the beginning. Scores were computed by taking an average of all items, after reversing items 1-5. An example item is, "I can trust the runner." This scale was reliable in this sample (α = .84).

Risk Avoidance. Risk avoidance was measured using the 10-item Risk Avoidance scale (see Appendix G; International Personality Item Pool, n.d.; α =.80). Participants were asked to describe themselves as they generally are now, not as they wish to be in the future using a five-point scale ranging from 1 (very inaccurate) to 5 (very accurate). All items were randomized in the study because all reverse-scored items were at the end. Scores were computed by taking an average of all items, with items 4-10 reverse scored. An example item is, "Would never make a high-risk investment." This scale was reliable in this sample (α = .74).

Video Game Use. A single-item question was used to assess how often participants played video games. Participants could respond using a five-point scale ranging from 1 (never) to 5 (daily). The item is, "How often do you play video games?"

Behavioral Trust. Betting behaviors, operationalized as loan amounts, on the part of the banker were used as the outcome variable to measure behavioral trust. There were three levels of trust that corresponded to the three loan amounts: small (low trust), medium (medium trust), and large (high trust). For example, when a participant chose to loan a small amount of money to the runner in round one, this indicated low trust.

Procedure

Experimental sessions were held in an on-campus laboratory. Participants were run individually. First, they were introduced to the Nao robot. The robot was located in the back room of the computer lab, crouching on a desk when participants arrived. The participants were told they were going to meet the other participant for the study, and then participants were walked into the back room to meet the robot. The experimenter tapped the robot on the head, which initiated the following speech and behavior. The robot stood up and became animated. Then the robot said the following, "Thanks for waking me up Sarah. Hi, I'm Rufus. It's nice to meet you. Time to get to work." Then the robot returned to the crouching position. Participants were then seated at a computer in the main room of the computer lab. After providing informed consent, participants completed surveys about demographics and video game use, then the CPRS, the original PTT, and the risk avoidance questionnaires. Next, participants completed an endowment earning task, which consisted of five, medium-difficulty, multiple choice math problems. Participants were told that based on their performance they would earn money towards the main task if they answered at least three out of five of the questions correctly. However, all participants earned $50 regardless of their performance. After the math task, participants completed the adapted PTT questionnaire. The experimenter read a backstory on Rufus aloud to participants (see Appendix H). Next, participants were provided with a PowerPoint training on Checkmate. After training, participants played two practice rounds of Checkmate with the robot, one round as the Banker and one round as the Runner. Participants were told that prior to coming in, they were randomly selected to play the Banker for the real session of five rounds and that the Nao robot was selected to play the Runner. Following practice, participants completed the perceived trustworthiness questionnaire. Round one lasted approximately three to five minutes. After the competition of the fifth round, participants were debriefed and paid for their time with a $30 gift card, and the money in their virtual wallet was paid to them in the form of cash.

III. RESULTS

Risk avoidance and video game use were not significantly correlated with any of the other measures, therefore they were not included for any of the following analyses. To test the hypothesis that propensity to trust automation would predict perceived trustworthiness (Hypothesis 1), Pearson's product-moment correlations were computed. Preliminary analyses showed all variables were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there were no outliers. A visual inspection of the scatterplots in each analysis each showed a linear relationship. A two-tailed, bivariate correlation ($\alpha = .05$) was computed between each of the three measures of propensity to trust automation and perceived trustworthiness.

CPRS and perceived trustworthiness were not significantly related ($r(53) = .08$, $p = .56$). There was a significant positive correlation between PTT and perceived trustworthiness ($r(53) = .48$, $p < .001$), and between the adapted version of PTT and perceived trustworthiness ($r(53) = .47$, $p < .001$). See Table 1 for all correlations, means, and standard deviations. Hypothesis 1 was partially supported. The PTT and the adapted PTT predicted perceived trustworthiness. However, the CPRS did not perceived trustworthiness.

To test the hypothesis that propensity to trust automation would predict behavioral trust (Hypothesis 2), Spearman's rank-order correlations were computed because the data were ordinal. A visual inspection of the scatterplots in each analysis each showed a monotonic relationship, and there were no outliers. A two-tailed, bivariate correlation ($\alpha = .05$) was computed between each of the three measures of propensity to trust automation and behavioral trust.

There was no significant correlation between CPRS and behavioral trust ($r_s(53) = .15$, $p = .27$), or between PTT and behavioral trust ($r_s(53) = .17$, $p = .23$). However, there was a significant positive correlation between the adapted version of PTT and behavioral trust ($r_s(53) = .30$, $p = .03$). Hypothesis 2 was partially supported. The adapted PTT predicted behavioral trust. However, the CPRS and the PTT did not predict behavioral trust.

Additionally, discriminant function analyses were conducted to examine if each measure of propensity to trust automation could individually predict whether participants would bet high, medium, or small loan amounts. Results indicated that the CPRS predicting participants' loan amounts was not significant, Wilk's $\lambda = .98$, χ^2 (2) = 1.32, $p = .52$, accounting for a non-significant 3% of the variance in betting behaviors. Approximately 29% of the participants in each group were correctly classified. The original PTT predicting loan amounts was not significant, Wilk's $\lambda = .96$, χ^2 (2) = 2.19, $p = .34$, accounting for a non-significant 4% of the variance in betting behaviors. Approximately 33% of the participants in each group were correctly classified. The adapted PTT predicting what category of loan amount participants chose was significant, Wilk's $\lambda = .88$, χ^2 (2) = 6.49, $p = .04$, accounting for a significant 12% of the variance in betting behaviors. Approximately 42% of the participants in each group were correctly classified. Again, Hypothesis 2 was partially supported. The CPRS and the original PTT did not predict whether participants would bet a high, medium, or low loan amount. However, the adapted PTT predicted participants loan amounts and accounted for 12% variance in betting behaviors.

To test whether the adapted propensity to trust in technology scale would better predict perceived trustworthiness compared to the other two propensity to trust automation scales (Hypothesis 3), a hierarchical multiple regression was computed. Preliminary assumption checking revealed there was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.85. There was homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There was one case with a studentized deleted residuals greater than ±3 standard deviations (*SD* = 3.25), no leverage values greater than 0.2, and no values for Cook's distance above 1. There assumption of normality was met, as assessed by Q-Q Plot. The data were analyzed with and without the outlier, and results were not significantly different; all cases were retained for the following analyses.

The first step of the hierarchical multiple regression regressing CPRS and PTT on perceived trustworthiness was significant, $R^2 = .24$, $F(2, 52) = 8.36$, $p = .001$, accounting for 24% of the variance in perceived trustworthiness (see Table 2). The second step included the adapted PTT measure, and this model was significant, $R^2 = .33$, $\Delta R^2 = .08$, $F(3, 51) = 8.21$, $p < .001$, accounting for 33% of the variance. The addition of the adapted PTT to the prediction of perceived trustworthiness accounted for an additional 9% of variance in perceived trustworthiness. Hypothesis 3 was supported. The adapted PTT accounted for a significant amount of additional variance in perceived trustworthiness, over and above that of the CPRS and PTT.

A combination of hierarchical multiple regression and discriminant function analysis was carried out to determine whether the addition of the adapted PTT improved the prediction of behavioral trust over and above CPRS and PTT alone (Hypothesis 4). Preliminary assumption testing revealed that there was linear relationship between the dependent variables in each loan category, as assessed by scatterplot, and no multicollinearity existed (the largest correlation was between PTT and adapted PTT; r = .52, $p < .001$). There were three univariate outliers but no multivariate outliers in the data, as assessed by boxplot and Mahalanobis distance ($p > .001$). The univariate outliers were retained in the following analyses. The data were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), except in the adapted PTT, small loan condition. As such, the data were not transformed. There was homogeneity of covariance matrices, as assessed by Box's M test ($p = .009$), and homogeneity of variances, as assessed by Levene's Test of Homogeneity of Variance ($p > .05$).

The first step of the hierarchical multiple regression regressed CPRS and PTT on behavioral trust, and it was not significant, $R^2 = .05$, $F(2, 52) = 1.28$, $p = .29$ (see Table 3). The combination of both CPRS and PTT accounted for a non-significant 5% of the variance in behavioral trust. The second step of the hierarchical multiple regression added adapted PTT as a predictor of perceived trustworthiness, and it was not significant,

$R^2 = .09$, $\Delta R^2 = .05$, $F(3, 51) = 1.71$, $p = .18$[2], accounting for 9% of the variance in behavioral trust. The addition of the adapted PTT to the prediction of behavioral trust accounted for an additional 4% of variance in behavioral trust, although this was not a significant increment.

A discriminant function analysis was conducted to predict group membership. The three groups were comprised of individuals who bet small loan amounts, those who bet medium loan amounts, and individuals who bet large loan amounts. Step one included both the CPRS and the PTT and was not significant, Wilk's $\lambda = .95$, χ^2 (4) = 2.54, $p = .64$, accounting for a non-significant 5% of the variance in betting behaviors. Just over 38% of the participants in each group were correctly classified. For step two, the adapted PTT was added to the model, Wilk's $\lambda = .86$, χ^2 (6) = 7.73, $p = .26$, accounting for a non-significant 12% of the variance. Almost half (45.5%) of the participants in each group were correctly classified. The addition of the adapted PTT into the model accounted for an extra 7% of the variance in behavioral trust, over and above that of the CPRS and PTT, but this was not significant. In the small loan category, 50% of the cases were predicted, in the medium loan category, 34.5% of the cases were predicted, and in the high loan category, 64.3% of the cases were predicted. Across all categories 45.5% of the original grouped cases were correctly classified. Hypothesis 4 was not supported. The adapted PTT did not predict behavior trust categories over and above the other two measures of propensity to trust automation.

---

[2] When outliers were removed, the first step became marginally significant ($p = .08$) and the second step became significant ($p = .04$).

IV. DISCUSSION

The purpose of this research was to compare three measures of propensity to trust automation and to determine whether these assessments are reliable and valid measures that help researchers better understand and predict both perceived trustworthiness and behavioral trust. Contrary to previous research (Merritt & Ilgen, 2008), the Complacency-Potential Rating Scale (CPRS) did not predict perceive trustworthiness. Also, the CPRS did not predict behavioral trust, which is similar to what other researchers have found (Merritt & Ilgen, 2008). One reason for these results could be that the CPRS is not a reliable scale (α = .67) or in previous research (α = .61; Merritt & Ilgen, 2008), as the commonly acceptable reliably coefficient is Cronbach's alpha of .70 or higher (Nunnally, 1978). Another possible reason that the CPRS did not predict either outcome in the present study could be that the measure is too broad, in that the items reference many different types of automation, such as ATMs, automated medical equipment, cruise control, etc. Each type of automation entails a certain level of risk associated with it and that risk may not be equal across the different types of automation. For example, if the cruise control on a car is faulty, the owner might be issued a ticket for going over the speed limit when she thought that she was going the legal speed limit. In contrast, if a person is undergoing open heart surgery and the surgery is augmented with a robotic arm for additional assistance, then the risk would be higher. If the automation fails in this second scenario, then the consequences could be fatal. Not all automated systems are created equally. As such, using a measure that is more consistent with the referent of interest—such as the adapted PTT that references "automated agent,"—increases reliability and predictive validity of the measure. Additionally, if people are given a more specific, single referent to focus on, then researchers can increase the precision of measurement (Lievens et al., 2008).

The Propensity to Trust in Technology (PTT) and the adapted PTT both predicted perceived trustworthiness. This is the first study to examine the use of the PTT to predict perceived trustworthiness of automation. Similar to previous research (Schneider et al., 2017), the PTT did not predict the participants' behavioral trust. Participants' general tendency to trust in technology did not influence the amount of money they chose to send to loan to the robot. However, the adapted PTT did predict behavioral trust. Participants disposition to trust in automated agents did predict the amount of money they loaned to the robot. To the author's knowledge, this is the first study that has found that propensity to trust automation predicts behavior. Although researchers in interpersonal trust (Evans & Revelle, 2008) have found that propensity to trust predicts behavioral trust, this phenomenon has not been found in the automation domain.

Adapting the PTT to use the term "automated agent" increased reliability from $\alpha = .76$ to $\alpha = .84$ and its capacity to predict both perceived trustworthiness and behavioral trust. The adapted PTT accounted for a significant amount of variance in perceived trustworthiness after controlling for the PTT and the CPRS. This supported the hypothesis that adapting a measure to make it more specific can better predict perceived trustworthiness. By using more specific language and decreasing ambiguity, the adapted

PTT was able to predict a significant amount of variance in perceived trustworthiness over and above the other measures. Additionally, in the second step of the hierarchical multiple regression, after controlling for the CPRS and PTT, the PTT was still significant. There was something unique about "automated agent" and "technology," in that both measures predicted perceived trustworthiness. Technology and automated agent were conceptualized by participants as separate referents and both predicted unique variance in perceived trustworthiness of the robot. However, this was not the case with behavioral trust. The adapted PTT was alone in predicting behavioral trust, but after controlling for the other two propensity to trust automation measures, the adapted PTT no longer significantly predicted behavioral trust. The two original measures shared variance in behavioral trust and adapting the PTT did not predict unique variance in betting behaviors. However, using correlational analysis, the adapted PTT accounted for 9% of the variance in behavioral trust. In the hierarchical multiple regression, even after controlling for the other two measures of propensity to trust automation, the adapted PTT still accounted for 9% of the variance in behavioral trust. Although the second step was not significant, it still highlights that the adapted PTT is an overall significant predictor of behavioral trust.

The theory of planned behavior states that behaviors are directly influenced by intentions (Ajzen, 1991). A measure of trust intentions was not included, nor the focus of this study. However, previous researchers have classified trust into three separate categories: trusting actions, trusting beliefs, and trusting intentions (Jones & Shah, 2016). Trusting actions are defined as an outcome that is the result of trusting in another, similar to behavioral trust. Trusting beliefs comprise positive expectations of the trustor that will influence her trust, such perceived trustworthiness in others. Trusting intention is the willingness a trustor has to be vulnerable to another. These three concepts are similar to those identified in Mayer et al.'s (1995) trust model. Future research should examine all three concepts and examine whether trust intentions mediates the propensity to trust and behavioral trust relationship in human-automation teams.

This study demonstrated that the more specific researchers are when referring to an automated system, the measures are more reliable, can account for more variance in perceived trustworthiness, and predict behavioral outcomes of trust. By decreasing the ambiguity of the referent in the propensity to trust automation scales and creating an adapted measure, the reliability and predictive validity was increased, compared to the original scales. This will help researchers better predict perceived trustworthiness of automated systems and how humans rely on automated systems in the future, in both personal use and job-related work.

A recent report by the McKinsey Global Institute (Manyika et al., 2017) estimated that by the year 2030, 60% of occupations around the globe will have at least a third of the job responsibilities automated. This means that job performance will be a combination of both human and automation work. Humans will have to learn how to work with automation, and automation will have to take into account how humans work to accomplish tasks together. Trust will play an important role in the human-automation relationship in that humans will have to trust the automation to perform jobs that were once completed by humans alone. It is also important to note that trust should appropriately calibrated to the automation (Lee & See, 2004). Over- or under-trusting an automated system can prove to be negligent or inefficient. For example, the 2009 crash of Air France Flight 447 was attributed to the pilots *over-trusting* the automation in the cockpit, which resulted in the death of all 228 passengers and crew aboard the aircraft (Todd & McConnell, 2011). On the other extreme, *under-trusting* in automation could result in employees engaging in unnecessary tasks and misusing their time. One example of this is a campus security company responsible for ensuring that multiple buildings are secure. If the doors are equipped with alarms, then the security guards should be notified of a breach. If the officers do not trust the automation, then they will be on a continuous rotation to manually check each door over the course of their shift. The time checking the doors could be better spent patrolling, responding to events, or writing reports. Companies may want individuals with appropriately calibrated levels of propensity to trust automation to occupy some of these positions with an automated component. Therefore, it is imperative to ensure employers are using the best measures they have during this selection process.

One limitation of this study was that there was only one type of automation that was used as the referent. Future research should use the adapted PTT with reference to other types of automation such as automated decision aids, automated cars, etc. This would expand upon this current research to test whether adapting the PTT scale to use the actual automated system as the referent can be replicated in other areas of automation. A second limitation that surfaced during pilot testing was that some participants did not understand what the term “automated agent” meant. A definition of an automated agent was added to the survey to help clarify any ambiguity. Future research should aim to include terms that 1) their samples understand, and 2) that their participants have experience with. A third limitation of this study was that it was mostly comprised of a convenience sample of college students. Future research should examine the reliability and validity of the adapted PTT in larger, more diverse samples. A fourth limitation was that this study had a small sample size. A priori power analysis indicated that a sample size of 84 was need for correlation analysis to detect a medium effect size. Therefore, there might not have been enough power in this study to detect a significant relationship between the CPRS and perceived trustworthiness, as well as the relationships between CPRS and behavioral trust, and PTT and behavioral trust. However, there was adequate power for both the hierarchical multiple regression and the discriminant function analysis. Future research should include a larger sample to provide more stable parameter estimates. Lastly, the order of the surveys was not counter-balanced, and participants could have experienced practice effects or item ordering effects. This could be addressed in future research by counterbalancing measures.

In conclusion, past research has been inconsistent in the ways that propensity to trust automation has been measured. Furthermore, the measures that are available have demonstrated low reliabilities. This study provided evidence that more specific measures of propensity to trust automation are more reliable and they are better predictors of perceived trustworthiness and behavioral trust. An adapted propensity to trust technology scale was the only significant predictor of both perceived trustworthiness of the automation and the trusting behaviors of participants. Researchers now have a more reliable and valid way to measure propensity to trust automation and researchers can incorporate this into future research. As more systems and machines in the workplace and everyday life become automated, it is important that researchers understand how individual factors, such as propensity to trust automation, influence trust and use of those systems.

V. REFERENCES

Adams, E. (2014). Failed airplane engine? Unconscious pilot? There's an app for that. Retrieved from http://www.popsci.com/xavion-ipad-app-can-make-emergency-airplane-landing-autopilot

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179-211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I., & Fishbein, M. (1973). Attitudinal and normative variables as predictors of specific behaviors. *Journal of Personality and Social Psychology*, *2*, 41-57. doi:10.1037/h0034440

Alarcon, G. M., Lyons, J. B., & Christensen, J. C. (2016). The effect of propensity to trust and familiarity on perceptions of trustworthiness over time. *Personality and Individual Differences*, *94*, 309-315. doi:10.1016/j.paid.2016.01.031

Alarcon, G. M., Lyons, J. B., Christensen, J. C., Klosterman, S. L., Bowers, M. A., Ryan, T. J., Jessup, S. A., & Wynne, K. T. (2017). The effect of propensity to trust and perceptions of trustworthiness on trust behaviors in dyads. *Behavioral Research Methods*. Advance online publication. doi:10.3758/s13428-017-0959-6

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*, 122-142. doi:10.1006/game.1995.1027

Bowling, N. A., & Burns, G. N. (2010). A comparison of work-specific and general personality measures as predictors of work and non-work criteria. *Personality and Individual Differences*, *49*, 95-101. doi:10.1016/j.paid.2010.03.009

Chancey, E. T., Bliss, J. P., Proaps, A. B., & Madhavan, P. (2015). The role of trust as a mediator between system characteristics and response behaviors. *Human Factors*, *57*, 947-958. doi:10.1177/0018720815582261

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*, 909-927. doi:10.1037/0021-9010.92.4.909

Costa, P. T., Jr., & McCrae, R. R. (1992). *The NEO PI–R professional manual*. Odessa, FL: Psychological Assessment Resources.

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, *13*, 123-139. doi:10.1177/001872676001300202

Erikson, E. H. (1950). Growth and crises of the “healthy personality”. In M. E. Senn (Eds.), *Symposium on the healthy personality* (pp. 91-146). Oxford, England: Josiah Macy, Jr. Foundation.

Evans, A. M., & Revelle, W. (2008). Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*, *42*, 1585–1593. doi:10.1016/j.jrp.2008.07.011

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BF03193146

Funder, D. C., & Ozer, D. J. (2010). *Pieces of the personality puzzle: Readings in theory and research*. New York, NY: W.W. Norton & Co.

Gill, H., Boies, K., Finegan, J. E., & McNally, J. (2005). Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust. *Journal of Business and Psychology*, *19*, 287-302. doi:10.1007/s10869-004-2229-8

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*, 407-434. doi:10.1177/0018720814547570

Huang, H., & Bashir, M. (2017). Personal influences on dynamic trust formation in human-agent interaction. *Proceedings of the International Conference on Human Agent Interaction*, *5*, 233-243. doi:10.1145/3125739.3125749

Jian, J., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*, 53-71. doi:10.1207/S15327566IJCE0401_04

Jones, S. L., & Shah, P. P. (2016). Diagnosing the locus of trust: A temporal perspective for trustor, trustee, and dyadic influences on perceived trustworthiness. *Journal of Applied Psychology*, *101*, 392-414. doi:10.1037/apl0000041

Langer, E. J. (1989). *Mindfulness*. Reading, MA: Addison-Wesley.

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*, 1243-1270.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50-80. doi:10.1518/hfes.46.1.50.30392

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, *93*, 268-279. doi:10.1037/0021-9010.93.2.268

Manyika, J., Lund, S. Chui, M., Bughin, J., Woetzel, J. Batra, P., Ko, R., & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation, *McKinsey Global Institute*, Retrieved from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/future%20of%20organizations/what%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/mgi%20jobs%20lost-jobs%20gained_report_december%202017.ashx

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*, 123-136. doi:10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*, 709-734. doi:10.2307/258792

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *The Academy of Management Review*, *23*, 473-490. doi:10.2307/259290

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*, 194-210. doi:10.1518/001872008X288574

Miller, A. P. (2018, July 26). Want less-biased decisions? Use algorithms. *Harvard Business Review*. Retrieved from https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Pop, V. L., Shrewsbury, A., & Durso, F. T. (2015). Individual differences in the calibration of trust in automation. *Human Factors*, *57*, 545-556. doi:10.1177/0018720814564422

Risk Avoidance Scale (n.d.) International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences. Retrieved from http://ipip.ori.org/newMPQKey.htm#Risk-AvoidanceS

Rosenberg, M. (1956). Misanthropy and political ideology. *American Sociological Review*, *21* (6), 690–695.

Rotter, J. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*, 651-665. doi:10.1111/j.1467-6494.1967.tb01454.x

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, *35*, 1-7. doi:10.1037/0003-066X.35.1.1

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*, 393-

404. doi:10.5465/AMR.1998.926617

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58,* 377–400. doi: 10.1177/0018720816634228

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*, 607. doi:10.1037/0021-9010.80.5.607

Schneider, T. R., Jessup, S. A., Stokes, C., Rivers, S., Lohani, M., & McCoy, M. (2017, May). *The influence of trust propensity on behavioral trust*. Poster session presented at the meeting of Association for Psychological Society, Boston, MA.

Schneider, T. R., Salovey, P., Apanovitch, A. M., Pizarro, J., McCarthy, D., Zullo, J., & Rothman, A. J. (2001). The effects of message framing and ethnic targeting on mammography use among low-income women. *Health Psychology, 20,* 256-266. doi:10.1037/0278-6133.20.4.256

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced “complacency”: Development of the Complacency-Potential Rating Scale. *The International Journal of Aviation Psychology*, *3*, 111-122. doi:10.1207/s15327108ijap0302_2

Tabachnick, B. G., Fidell, L. S. (2007) *Using multivariate statistics*. Boston, MA: Pearson.

Technology. (n.d.). Retrieved from https://www.merriam-webster.com/dictionary/technology

Todd, B., & McConnell, D. (2011). Autopilots may dull skills of pilots, committee says, *CNN*, Retrieved from http://www.cnn.com/2011/TRAVEL/09/01/airlines.autopilot/

Van Lange, P. A. M. (2015). Generalized trust: Four lessons from genetics and culture. *Current Directions in Psychological Science*, *24*, 71-76. doi:10.1177/0963721414552473

Wiggins, J. S. (1973). *Personality and prediction: principles of personality assessment*. Boston, MA: Addison-Wesley.

Table 1

Descriptive Statistics and Correlations Among the Demographic and Scale Variables

	Mean	*SD*	1	2	3	4	5	6	7	8
1. Age	24.11	5.36								
2. Gender			-.24							
3. Video Game Use	2.80	1.38	-.10	-.47**						
4. Risk Avoidance	3.07	.64	.08	.10	.03	(.74)				
5. CPRS	3.79	.42	.11	-.32*	.10	-.07	(.67)			
6. PTT (original)	4.00	.49	.10	-.12	.12	-.04	.43**	(.76)		
7. PTT (adapted)	3.58	.59	.16	-.12	-.07	.01	.33*	.52**	(.84)	
8. Perceived Trustworthiness	4.84	.81	.04	-.07	-.07	-.04	.08	.48**	.47**	(.83)
9. Behavioral Trust	1.04	.69	.09	-.07	.21	.01	.15	.17	.30*	.36**

Note. N = 55 (except for age N = 54). * p < .05. ** p < .01. Cronbach's alphas are reported in parentheses on the diagonal. Gender was coded: Males = 0, Females = 1. For the Risk Avoidance, Complacency-Potential Rating Scale (CPRS), and Propensity to Trust in Technology (PTT) items were measured on 5-point response scales; Perceived Trustworthiness items were measured on 7-point response scales. Behavioral Trust was measured using three levels: Small = 0, Medium = 1, Large = 2. All variables were analyzed using Pearson-Product Moment correlation, except behavioral trust, which was analyzed with Spearman Rank-Order correlation.

Table 2

Hierarchical Multiple Regression Analysis Predicting Perceived Trustworthiness.

	Perceived Trustworthiness			
	Step 1		Step 2	
Predictor	B	β	B	β
Constant	2.38*		2.09*	
CPRS	-.28	-.15	-.37	-.20
PTT	.89**	.54	.63*	.38
Adapted PTT			.46*	.34
R^2	.24		.33	
F	8.36*		8.21*	
ΔR^2			.08	
ΔF			6.22*	

Note. N = 55. * p < .05, ** p < .001. CPRS = Complacency-Potential Rating Scale; PTT = Propensity to Trust in Technology.

Table 3

Hierarchical Multiple Regression Analysis Predicting Behavioral Trust.

	Behavioral Trust			
	Step 1		Step 2	
Predictor	B	β	B	β
Constant	-.44		.62	
CPRS	.15	.09	.09	.05
PTT	.23	.16	.07	.05
Adapted PTT			.29	.25
R^2	.05		.09	
F	1.28		1.71	
ΔR^2			.05	
ΔF			2.50	

Note. N = 55. * p < .05, ** p < .001. CPRS = Complacency-Potential Rating Scale; PTT = Propensity to Trust in Technology.

Appendix A

Complacency-Potential Rating Scale
(Singh, Molloy, & Parasuraman, 1993)

Instructions: Please indicate your agreement or disagreement with each of these statements using the 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree).

1. I think that automated devices used in medicine, such as CT scans and ultrasound, provide very reliable medical diagnosis.

2. Automated devices in medicine save time and money in the diagnosis and treatment of a disease.

3. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because it is more reliable and safer than manual surgery.

4. Automated systems used in modern aircraft, such as the automatic landing system, have made air journeys safer.

5. ATMs provide a safeguard against the inappropriate use of an individual's bank account against dishonest people.

6. Automated devices used in aviation and banking have made work easier for both employees and customers.

7. Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed trap in case the automatic control is not working properly. (R)

8. Manually sorting through card catalogues is more reliable than computer-aided searches for finding items in a library. (R)

9. I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.

10. Bank transactions have become safer with the introduction of computer

technology for the transfer of funds.

11. I feel safer depositing my money at an ATM than with a human teller.

12. I have to record an important TV program for a class assignment. To ensure that the correct program is recorded, I would use the automatic programming facility on my DVR rather than manual recording.

13. People save time by using automatic teller machines (ATMs) rather than a bank teller in making transactions.

14. I often use automated devices.

Appendix B

Propensity to Trust in Technology
(Schneider et al., 2017)

Instructions: For the below listed items, please read each statement carefully. Using the 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree), select the answer that most accurately describes your feelings.

1. Generally, I trust technology.

2. Technology helps me solve many problems.

3. I think it’s a good idea to rely on technology for help.

4. I don’t trust the information I get from technology. (R)

5. Technology is reliable.

6. I rely on technology.

Appendix C

NAO Robot

Appendix D

Adapted Propensity to Trust in Technology
(Schneider et al., 2017)

Instructions: For the below listed items, please read each statement carefully. Using the 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree), select the answer that most accurately describes your feelings.

1. Generally, I trust automated agents.
2. Automated agents help me solve many problems.
3. I think it's a good idea to rely on automated agents for help.
4. I don't trust the information I get from automated agents. (R)
5. Automated agents are reliable.
6. I rely on automated agents.

Appendix E

Trust in Automated Systems
(Jian, Bisantz, & Drury, 2000)

<ins>Instructions:</ins> Below is a list of statements for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating a machine. Please select the option which best describes your feeling or your impression using the 7-point scale ranging from 1 (not at all) to 7 (extremely).

1. The system is deceptive. (R)
2. The system behaves in an underhanded manner. (R)
3. I am suspicious of the system's intent, action, and outputs. (R)
4. I am wary of the system. (R)
5. The system's actions will have a harmful or injurious outcome. (R)
6. I am confident in the system.
7. The system provides security.
8. The system has integrity.
9. The system is dependable.
10. The system is reliable.
11. I can trust the system.
12. I am familiar with the system.

Appendix F

Adapted Trust in Automated Systems
(Jian, Bisantz, & Drury, 2000)

Instructions: Below is a list of statements for evaluating trust. There are several items for you to rate intensity of your feeling of trust, or your impression of the runner while engaging in a task. Please select the option which describes your feeling or your impression using the 7-point scale ranging from 1 (not at all) to 7 (extremely).

1. The runner is deceptive. (R)
2. The runner behaves in an underhanded manner. (R)
3. I am wary of the runner. (R)
4. The runner's actions will have a harmful or injurious outcome. (R)
5. I am confident in the runner.
6. The runner provides security.
7. The runner has integrity.
8. The runner is dependable.
9. The runner is reliable.
10. I can trust the runner.
11. I am familiar with the runner.

Appendix G

Risk-Avoidance
(International Personality Item Pool)

<u>Instructions:</u> Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

1. Would never go hang gliding or bungee jumping.
2. Would never make a high-risk investment.
3. Avoid dangerous situations.
4. Seek danger. (R)
5. Am willing to try anything once. (R)
6. Do dangerous things. (R)
7. Enjoy being reckless. (R)
8. Seek adventure. (R)
9. Take risks. (R)
10. Do crazy things. (R)

Appendix H

Backstory for Rufus

“The military currently integrates automation into dangerous scenarios alongside humans. Automation is useful in high-risk scenarios, such as disabling explosive devices, navigating unmanned aerial vehicles (UAVs), and carrying heavy equipment. However, automation is expensive and takes time to develop. As such, the military is testing automated robots containing self-preservation algorithms. This means the military is creating robots that should be able to make decisions to protect themselves, as well as other humans around them. If a situation is too dangerous, the robot should take proper precautions to minimize damages to itself. The current study uses the same algorithms to aid the robot’s decision-making when teamed with another human in a maze-running task. Keep in mind that Rufus the robot may act self-interested, meaning he may prioritize himself over you.”